09/674588 PCT/JP00/01282

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: １９９９年　３月　４日

出願番号
Application Number: 平成１１年特許願第０５７４６７号

出願人
Applicant (s): ソニー株式会社

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

２０００年　２月　４日

特許庁長官
Commissioner,
Patent Office 近藤隆彦

出証番号　出証特２０００－３００３２７７

【書類名】　特許願

【整理番号】　9801154103

【提出日】　平成11年 3月 4日

【あて先】　特許庁長官殿

【国際特許分類】　G06K 9/62

G10L 3/00

【発明者】

【住所又は居所】　東京都品川区北品川６丁目７番３５号　ソニー株式会社内

【氏名】　包　洪長

【特許出願人】

【識別番号】　000002185

【氏名又は名称】　ソニー株式会社

【代表者】　出井　伸之

【代理人】

【識別番号】　100082131

【弁理士】

【氏名又は名称】　稲本　義雄

【電話番号】　03-3369-6479

【手数料の表示】

【予納台帳番号】　032089

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9708842

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　パターン認識装置および方法、並びに提供媒体

【特許請求の範囲】

【請求項１】　入力されるデータの特徴分布を、所定数のモデルのうちのいずれかに分類するパターン認識装置において、

前記入力されるデータのパターンを特徴分布として抽出する抽出手段と、

前記所定数のモデルを記憶する記憶手段と、

前記抽出手段が抽出した特徴分布を、前記所定数のモデルのうちのいずれかに分類する分類手段と、

前記データが入力される直前に入力されたノイズに基づいて、前記データが存在しない状態に対応する前記モデルを生成し、前記記憶手段に記憶されている対応するものを更新する生成手段と

を含むことを特徴とするパターン認識装置。

【請求項２】　前記データが存在しない状態の特徴分布、および、前記データが存在しない状態に対応する前記モデルの確率分布が正規分布である場合、前期生成手段は、前記データが存在しない状態に対応する前記モデルの期待値を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する期待値の平均として生成し、前記データが存在しない状態に対応する前記モデルの分散を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する分散の平均として生成する

ことを特徴とする請求項１に記載のパターン認識装置。

【請求項３】　前記データが存在しない状態の特徴分布、および、前記データが存在しない状態に対応する前記モデルの確率分布が正規分布である場合、前期生成手段は、前記データが存在しない状態に対応する前記モデルの期待値および分散を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する期待値の平均を用いて生成する

ことを特徴とする請求項１に記載のパターン認識装置。

【請求項４】　前記データが存在しない状態の特徴分布、および、前記データが存在しない状態に対応する前記モデルの確率分布が正規分布である場合、前

期生成手段は、前記データが存在しない状態に対応する前記モデルの確率分布を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する統計量の線形結合に基づいて生成する

ことを特徴とする請求項１に記載のパターン認識装置。

【請求項５】　前記データが存在しない状態の特徴分布、および、前記データが存在しない状態に対応する前記モデルの確率分布が正規分布である場合、前期生成手段は、前記データが存在しない状態に対応する前記モデルの確率分布を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する統計母集団の和に基づいて生成する

ことを特徴とする請求項１に記載のパターン認識装置。

【請求項６】　入力されるデータの特徴分布を、所定数のモデルのうちのいずれかに分類するパターン認識装置のパターン認識方法において、

前記入力されるデータのパターンを特徴分布として抽出する抽出ステップと、

前記所定数のモデルを記憶する記憶ステップと、

前記抽出ステップで抽出した特徴分布を、前記所定数のモデルのうちのいずれかに分類する分類ステップと、

前記データが入力される直前に入力されたノイズに基づいて、前記データが存在しない状態に対応する前記モデルを生成し、前記記憶ステップで記憶された対応するものを更新する生成ステップと

を含むことを特徴とするパターン認識方法。

【請求項７】　入力されるデータの特徴分布を、所定数のモデルのうちのいずれかに分類するパターン認識装置に、

前記入力されるデータのパターンを特徴分布として抽出する抽出ステップと、

前記所定数のモデルを記憶する記憶ステップと、

前記抽出ステップで抽出した特徴分布を、前記所定数のモデルのうちのいずれかに分類する分類ステップと、

前記データが入力される直前に入力されたノイズに基づいて、前記データが存在しない状態に対応する前記モデルを生成し、前記記憶ステップで記憶された対応するものを更新する生成ステップと

を含む処理を実行させるコンピュータが読み取り可能なプログラムを提供することを特徴とする提供媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、パターン認識装置および方法、並びに提供媒体に関し、特に、ノイズ環境下において発話された単語を認識するパターン認識装置および方法、並びに提供媒体に関する。

【０００２】

【従来の技術】

従来より、ノイズ環境下において発話された単語を識別する方法が考案されており、その代表的な方法としては、PMC(Parallel Model Combination)法、SS/NSS(Spectral Subtraction/Nonlinear Spectral Subtraction)法、SFE(Stochastic Feature Extraction)法等が知られている。

【０００３】

上述したいずれの方法においても、発話音声に環境ノイズが混在している音声データの特徴量が抽出され、その特徴量が、予め登録されている複数の単語に対応する音響モデルのうちのいずれに最も適合するかが判定されて、最も適合する音響モデルに対応する単語が認識結果として出力される。

【０００４】

上述した方法の特徴は、以下の通りである。すなわち、PMC法は、環境ノイズの情報を直接的に音響モデルに取り込んでいるので認識性能は良いが、計算コストが高い（高度な演算を必要とするので、装置の規模が大型化する、処理に要する時間が長い等）。SS/NSS法は、音声データの特徴量を抽出する段階において、環境ノイズを除去している。したがって、PMC法よりも計算コストが低く、現在、多く用いられている方法である。なお、SS/NSS法は、音声データの特徴量をベクトルとして抽出する。SFE法は、SS/NSS法と同様に、ミックス信号の特徴量を抽出する段階において、環境ノイズを除去するが、特徴量を確率分布として抽出する。

【０００５】

【発明が解決しようとする課題】

ところで、SFE法においては、音声認識の段階で環境ノイズが直接的に反映されていない、すなわち、環境ノイズの情報が直接的に無音音響モデルに取り込まれていないので、認識性能が劣る課題があった。

【０００６】

また、環境ノイズの情報が直接的に無音音響モデルに取り込まれていないことに起因して、音声認識が開始された時点から発話が開始されるまでの時間が長くなるにつれて認識性能が低下する課題があった。

【０００７】

本発明はこのような状況に鑑みてなされたものであり、環境ノイズの情報を用いて無音音響モデルを補正することにより、音声認識が開始された時から発話が開始される時までの時間が長くなるに伴って認識性能が低下することを抑止するようにするものである。

【０００８】

【課題を解決するための手段】

請求項１に記載のパターン認識装置は、入力されるデータのパターンを特徴分布として抽出する抽出手段と、所定数のモデルを記憶する記憶手段と、抽出手段が抽出した特徴分布を、所定数のモデルのうちのいずれかに分類する分類手段と、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルを生成し、記憶手段に記憶されている対応するものを更新する生成手段とを含むことを特徴とする。

【０００９】

請求項６に記載のパターン認識方法は、入力されるデータのパターンを特徴分布として抽出する抽出ステップと、所定数のモデルを記憶する記憶ステップと、抽出ステップで抽出した特徴分布を、所定数のモデルのうちのいずれかに分類する分類ステップと、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルを生成し、記憶ステップで記憶された対応するものを更新する生成ステップとを含むことを特徴とする。

【００１０】

請求項７に記載の提供媒体は、入力されるデータのパターンを特徴分布として抽出する抽出ステップと、所定数のモデルを記憶する記憶ステップと、抽出ステップで抽出した特徴分布を、所定数のモデルのうちのいずれかに分類する分類ステップと、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルを生成し、記憶ステップで記憶された対応するものを更新する生成ステップとをを含む処理をパターン認識装置に実行させるコンピュータが読み取り可能なプログラムを提供することを特徴とする。

【００１１】

請求項１に記載のパターン認識装置、請求項６に記載のパターン認識方法、および請求項７に記載の提供媒体においては、入力されるデータのパターンが特徴分布として抽出され、所定数のモデルが記憶されて、抽出された特徴分布が、所定数のモデルのうちのいずれかに分類される。また、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルが生成されて記憶されている対応するものが更新される。

【００１２】

【発明の実施の形態】

本発明を適用した音声認識装置の構成例について、図１を参照して説明する。この音声認識装置において、マイクロフォン１は、認識対象である発話音声を、環境ノイズとともに集音し、フレーム化部２に出力する。フレーム化部２は、マイクロフォン１から入力される音声データを、所定の時間間隔（例えば、１０ｍｓ）で取り出し、その取り出したデータを、１フレームのデータとして出力する。フレーム化部２が出力する１フレーム単位の音声データは、そのフレームを構成する時系列の音声データそれぞれをコンポーネントとする観測ベクトルａとして、ノイズ観測区間抽出部３、および特徴抽出部５に供給される。

【００１３】

ここで、以下、適宜、第ｔフレームの音声データである観測ベクトルを、ａ（ｔ）と表す。

【００１４】

ノイズ観測区間抽出部３は、フレーム化部２から入力されるフレーム化された音声データを所定の時間（Ｍフレーム分以上）だけバッファリングしており、図２に示すように、発話スイッチ４がオンとされたタイミングt_bからＭフレーム分だけ以前のタイミングt_aまでをノイズ観測区間Ｔｎとし、ノイズ観測区間Ｔｎにおけるフレーム分の観測ベクトルａを抽出して特徴抽出部５、および無音音響モデル補正部７に出力する。

【００１５】

発話スイッチ４は、ユーザが発話を開始するときにユーザ自身がオンとするスイッチである。したがって、発話スイッチ４がオンとされたタイミングt_b以前（ノイズ観測区間Ｔｎ）の音声データには、発話音声は含まれず、環境ノイズだけが存在する。また、発話スイッチ４がオンとされたタイミングt_bから発話スイッチ４がオフとされるタイミングt_dまでは、音声認識区間とされて、その区間の音声データが音声認識の対象とされる。

【００１６】

特徴抽出部５は、ノイズ観測区間抽出部３から入力されるノイズ観測区間Ｔｎの環境ノイズだけが存在する音声データに基づいて、フレーム化部２から入力される、タイミングt_b以降の音声認識区間の観測ベクトルａから環境ノイズ成分を除去して、その特徴量を抽出する。すなわち、特徴抽出部５は、例えば、観測ベクトルａとしての真（環境ノイズが除去された）の音声データをフーリエ変換し、そのパワースペクトラムを求め、そのパワースペクトラムの各周波数成分をコンポーネントとする特徴ベクトルｙを算出する。なお、パワースペクトラムの算出方法は、フーリエ変換によるものに限定されるものではない。すなわち、パワースペクトラムは、その他、例えば、いわゆるフィルタバンク法などによって求めることも可能である。

【００１７】

さらに、特徴抽出部５は、観測ベクトルａとしての音声データに含まれる音声を、その特徴量の空間（特徴ベクトル空間）に写像したときに得られる、その特徴ベクトル空間上の分布を表すパラメータ（以下、特徴分布パラメータと記述す

る）Ｚを、算出した特徴ベクトルｙに基づいて算出し、音声認識部６に供給する。

【００１８】

図３は、図１の特徴抽出部５の詳細な構成例を示している。フレーム化部２から入力される観測ベクトルａは、特徴抽出部５において、パワースペクトラム分析部１１に供給される。パワースペクトラム分析部１１では、観測ベクトルａが、例えば、FFT（高速フーリエ変換）アルゴリズムによってフーリエ変換され、これにより、音声の特徴量であるパワースペクトラムが、特徴ベクトルとして抽出される。なお、ここでは、１フレームの音声データとしての観測ベクトルａが、Ｄ個のコンポーネントからなる特徴ベクトル（Ｄ次元の特徴ベクトル）に変換されるものとする。

【００１９】

ここで、第ｔフレームの観測ベクトルａ（ｔ）から得られる特徴ベクトルをｙ（ｔ）と表す。また、特徴ベクトルｙ（ｔ）のうち、真の音声のスペクトル成分をｘ（ｔ）と、環境ノイズのスペクトル成分をｕ（ｔ）と表す。この場合、真の音声のスペクトル成分ｘ（ｔ）は、次式（１）で表される。

【数１】

$$x(t)=y(t)-u(t) \qquad \cdots (1)$$

ただし、ここでは、環境ノイズが不規則な特性を有し、また、観測ベクトルａ（ｔ）としての音声データは、真の音声成分に環境ノイズを加算したものであると仮定されている。

【００２０】

一方、ノイズ観測区間抽出部３から入力される音声データ（環境ノイズ）は、特徴検出部５において、ノイズ特性算出部１３に入力される。ノイズ特性算出部１３では、ノイズ観測区間Ｔｎにおける環境ノイズの特性が求められる。

【００２１】

すなわち、ここでは、音声認識区間における環境ノイズのパワースペクトラムｕ（ｔ）の分布が、その音声認識区間の直前のノイズ観測区間Ｔｎにおける環境

ノイズと同一であるとし、さらに、その分布が正規分布であるとして、ノイズ特性算出部１３において、環境ノイズの平均値（平均ベクトル）と分散（分散マトリクス）が求められる。

【００２２】

なお、平均ベクトルμ'と分散マトリクスΣ'は、次式（２），（３）にしたがって求めることができる。

【数２】

$$\mu'(i)=\frac{1}{M}\sum_{t=1}^{M} y(t)(i)$$

$$\Sigma'(i,j)=\frac{1}{M}\sum_{t=1}^{M}(y(t)(i)-\mu'(i))(y(t)(j)-\mu'(j)) \quad \cdots(2)$$

ただし、μ'（ｉ）は、平均ベクトルμ'のｉ番目のコンポーネントを表す（ｉ＝１，２，・・・，Ｄ）。また、ｙ（ｔ）（ｉ）は、第ｔフレームの特徴ベクトルのｉ番目のコンポーネントを表す。さらに、Σ'（ｉ，ｊ）は、分散マトリクスΣ'の、第ｉ行、第ｊ列のコンポーネントを表す（ｊ＝１，２，・・・，Ｄ）。

【００２３】

ここで、計算量の低減のために、環境ノイズについては、特徴ベクトルｙの各コンポーネントが、互いに無相関であると仮定する。この場合、次式に示すように、分散マトリクスΣ'は、対角成分以外は０となる。

【数３】

$$\Sigma'(i,j)=0,\ i\neq j \quad \cdots(3)$$

【００２４】

ノイズ特性算出部１３では、以上のようにして、環境ノイズの特性としての平均ベクトルμ'および平均値Σ'が求められ、特徴分布パラメータ算出部１２に供給される。

【００２５】

一方、パワースペクトラム分析部１１の出力、すなわち、環境ノイズを含む発話音声の特徴ベクトルｙは、特徴分布パラメータ算出部１２に供給される。特徴分布パラメータ算出部１２では、パワースペクトラム分析部１１からの特徴ベクトルｙ、およびノイズ特性算出部１３からの環境ノイズの特性に基づいて、真の音声のパワースペクトラムの分布（推定値の分布）を表す特徴分布パラメータが算出される。

【００２６】

すなわち、特徴分布パラメータ算出部１２では、真の音声のパワースペクトラムの分布が正規分布であるとして、その平均ベクトルξと分散マトリクスΨが、特徴分布パラメータとして、次式（４）乃至（７）にしたがって計算される。

【数４】

$$\begin{aligned}\xi(t)(i) &= E[x(t)(i)] \\ &= E[y(t)(i)-u(t)(i)] \\ &= \int_0^{y(t)(i)} (y(t)(i)-u(t)(i)) \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i))du(t)(i)} du(t)(i) \\ &= \frac{y(t)(i)\int_0^{y(t)(i)} P(u(t)(i))du(t)(i) - \int_0^{y(t)(i)} u(t)(i)P(u(t)(i))du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i))du(t)(i)} \\ &= y(t)(i) - \frac{\int_0^{y(t)(i)} u(t)(i)P(u(t)(i))du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i))du(t)(i)} \qquad \cdots (4)\end{aligned}$$

【数５】

i=jのとき

$$\Psi(t)(i,j)=V[x(t)(i)]$$
$$=E[(x(t)(i))^2]-(E[x(t)(i)])^2$$
$$(=E[(x(t)(i))^2]-(\xi(t)(i))^2)$$

i≠jのとき

$$\Psi(t)(i,j)=0 \qquad \cdots(5)$$

【数６】

$$E[(x(t)(i))^2]=E[(y(t)(i)-u(t)(i))^2]$$

$$=\int_0^{y(t)(i)}(y(t)(i)-u(t)(i))^2\frac{P(u(t)(i))}{\int_0^{y(t)(i)}P(u(t)(i))du(t)(i)}du(t)(i)$$

$$=\frac{1}{\int_0^{y(t)(i)}P(u(t)(i))du(t)(i)}\times\left\{(y(t)(i))^2\int_0^{y(t)(i)}P(u(t)(i))du(t)(i)\right.$$

$$-2y(t)(i)\int_0^{y(t)(i)}u(t)(i)P(u(t)(i))du(t)(i)$$

$$\left.+\int_0^{y(t)(i)}(u(t)(i))^2P(u(t)(i))du(t)(i)\right\}$$

$$=(y(t)(i))^2-2y(t)(i)\frac{\int_0^{y(t)(i)}u(t)(i)P(u(t)(i))du(t)(i)}{\int_0^{y(t)(i)}P(u(t)(i))du(t)(i)}$$

$$+\frac{\int_0^{y(t)(i)}(u(t)(i))^2P(u(t)(i))du(t)(i)}{\int_0^{y(t)(i)}P(u(t)(i))du(t)(i)} \qquad \cdots(6)$$

【数７】

$$P(u(t)(i))=\frac{1}{\sqrt{2\pi\,\Sigma'(i,i)}}\,e^{-\frac{1}{2\Sigma'(i,i)}(u(t)(i)-\mu'(i))^2} \quad \cdots (7)$$

【００２７】

ここで、ξ（ｔ）（ｉ）は、第ｔフレームにおける平均ベクトルξ（ｔ）のｉ番目のコンポーネントを表す。また、Ｅ［］は、［］内の平均値を意味する。ｘ（ｔ）（ｉ）は、第ｔフレームにおける真の音声のパワースペクトラムｘ（ｔ）のｉ番目のコンポーネントを表す。さらに、ｕ（ｔ）（ｉ）は、第ｔフレームにおける環境ノイズのパワースペクトラムのｉ番目のコンポーネントを表し、Ｐ（ｕ（ｔ）（ｉ））は、第ｔフレームにおける環境ノイズのパワースペクトラムのｉ番目のコンポーネントがｕ（ｔ）（ｉ）である確率を表す。ここでは、環境ノイズの分布として正規分布を仮定しているので、Ｐ（ｕ（ｔ）（ｉ））は、式（７）に示したように表される。

【００２８】

また、Ψ（ｔ）（ｉ，ｊ）は、第ｔフレームにおける分散Ψ（ｔ）の、第ｉ行、第ｊ列のコンポーネントを表す。さらに、Ｖ［］は、［］内の分散を表す。

【００２９】

特徴分布パラメータ算出部１２では、以上のようにして、各フレームごとに、平均ベクトルξおよび分散マトリクスΨが、真の音声の特徴ベクトル空間上での分布（ここでは、真の音声の特徴ベクトル空間上での分布が正規分布であると仮定した場合の、その分布）を表す特徴分布パラメータとして求められる。

【００３０】

その後、音声認識区間の各フレームにおいて求めた特徴分布パラメータは、音声認識部６に出力される。すなわち、いま、音声認識区間がＴフレームであったとし、そのＴフレームそれぞれにおいて求められた特徴分布パラメータを、ｚ（ｔ）＝｛ξ（ｔ），Ψ（ｔ）｝（ｔ＝１，２，・・・，Ｔ）と表すと、特徴分布パラメータ算出部１２は、特徴分布パラメータ（系列）Ｚ＝｛ｚ（１），ｚ（２

），・・・，ｚ（Ｔ）｝を、音声認識部６に供給する。

【００３１】

図１に戻る。音声認識部６は、特徴抽出部５から入力される特徴分布パラメータＺを、所定数Ｋの音響モデルと１個の無音音響モデルのうちのいずれかに分類し、その分類結果を、入力された音声の認識結果として出力する。すなわち、音声認識部６は、例えば、無音区間に対応する識別関数（特徴パラメータＺが無音音響モデルに分類されるかを識別するための関数）と、所定数Ｋの単語それぞれに対応する識別関数（特徴パラメータＺがいずれの音響モデルに分類されるかを識別するための関数）とを記憶しており、各音響モデルの識別関数の値を、特徴抽出部５からの特徴分布パラメータＺを引数として計算する。そして、その関数値が最大である音響モデル（単語、または無音区間）が認識結果として出力される。

【００３２】

図４は、図１の音声認識部６の詳細な構成例を示している。特徴抽出部５の特徴分布パラメータ算出部１２から入力される特徴分布パラメータＺは、識別関数演算部２１－１乃至２１－ｋ、および識別関数演算部２１－ｓ、に供給される。識別関数演算部２１－ｋ（ｋ＝１，２，・・・，Ｋ）は、Ｋ個の音響モデルのうちのｋ番目に対応する単語を識別するための識別関数G_k（Ｚ）を記憶しており、特徴抽出部５からの特徴分布パラメータＺを引数として、識別関数G_k（Ｚ）を演算する。識別関数演算部２１－ｓは、無音音響モデルに対応する無音区間を識別するための識別関数G_s（Ｚ）を記憶しており、特徴抽出部５からの特徴分布パラメータＺを引数として、識別関数G_s（Ｚ）を演算する。

【００３３】

なお、音声認識部６では、例えば、HMM(Hidden Markov Model)法を用いて、クラスとしての単語または無音区間の識別（認識）が行われる。

【００３４】

HMM法について、図５を参照して説明する。同図において、HMMは、Ｈ個の状態q_1乃至q_Hを有しており、状態の遷移は、自身への遷移と、右隣の状態への遷移のみが許されている。また、初期状態は、最も左の状態q_1とされ、最終状態は

、最も右の状態q_Hとされており、最終状態q_Hからの状態遷移は禁止されている。このように、自身よりも左にある状態への遷移のないモデルは、left-to-rightモデルと呼ばれ、音声認識では、一般に、left-to-rightモデルが用いられる。

【００３５】

いま、HMMのｋクラスを識別するためのモデルを、ｋクラスモデルというとすると、ｋクラスモデルは、例えば、最初に状態q_hにいる確率（初期状態確率）$\pi_k(q_h)$、ある時刻（フレーム）ｔにおいて、状態q_iにいて、次の時刻ｔ＋１において、状態q_jに状態遷移する確率（遷移確率）$a_k(q_i, q_j)$、および状態q_iから状態遷移が生じるときに、その状態q_iが、特徴ベクトルＯを出力する確率（出力確率）$b_k(q_i)(O)$によって規定される（ｈ＝１，２，・・・，Ｈ）。

【００３６】

そして、ある特徴ベクトル系列O_1，O_2，・・・が与えられた場合、例えば、そのような特徴ベクトル系列が観測される確率（観測確率）が最も高いモデルのクラスが、その特徴ベクトル系列の認識結果とされる。

【００３７】

ここでは、この観測確率が、識別関数$G_k(Z)$によって求められる。すなわち、識別関数$G_k(Z)$は、特徴分布パラメータ（系列）$Z=\{z_1, z_2, \cdots, z_T\}$に対する最適状態系列（最適な状態の遷移のしていき方）において、そのような特徴分布パラメータ（系列）$Z=\{z_1, z_2, \cdots, z_T\}$が観測される確率を求めるものとして、次式（８）で与えられる。

【数８】

$$g_k(Z)=\max_{q_1,q_2,\cdots,q_T} \pi_k(q_1)\cdot b_k'(q_1)(z_1)\cdot a_k(q_1,q_2)\cdot b_k'(q_2)(z_2) \cdots a_k(q_{T-1},q_T)\cdot b_k'(q_T)(z_T) \quad \cdots (8)$$

【００３８】

ここで、$b_k'(q_i)(z_j)$は、出力がz_jで表される分布であるときの出力確率を表す。状態遷移時に各特徴ベクトルを出力する確率である出力確率b_k（

ｓ）（O_t）には、ここでは、例えば、特徴ベクトル空間上のコンポーネントに相関がないものとして、正規分布関数が用いられている。この場合、入力がz_tで表される分布であるとき、出力確率b_k'（ｓ）（z_t）は、平均ベクトルμ_k（ｓ）と分散マトリクスΣ_k（ｓ）とによって規定される確率密度関数P_k^m（ｓ）（ｘ）、および第ｔフレームの特徴ベクトル（ここでは、パワースペクトラム）ｘの分布を表す確率密度関数P^f（ｔ）（ｘ）を用いて、次式（９）により求めることができる。

【数９】

$$b_k'(s)(z_t)=\int P^f(t)(x)P_k^m(s)(x)dx$$

$$=\prod_{i=1}^{D} P(s)(i)(\xi(t)(i),\Psi(t)(i,i))$$

$$k=1,2,\cdots,K : s=q_1,q_2\cdots,q_T : T=1,2\cdots,T \quad \cdots(9)$$

ただし、式（９）における積分の積分区間は、Ｄ次元の特徴ベクトル空間（ここでは、パワースペクトラム空間）の全体である。

【００３９】

また、式（９）において、Ｐ（ｓ）（ｉ）（ξ（ｔ）（ｉ），Ψ（ｔ）（ｉ，ｉ））は、次式（１０）で表される。

【数１０】

$$P(s)(i)(\xi(t)(i),\Psi(t)(i,i))$$

$$=\frac{1}{\sqrt{2\pi(\Sigma_k(s)(i,i)+\Psi(t)(i,i))}}\,e^{-\frac{(\mu_k(s)(i)-\xi(t)(i))^2}{2(\Sigma_k(s)(i,i)+\Psi(t)(i,i))}} \quad \cdots(10)$$

ただし、μ_k（ｓ）（ｉ）は、平均ベクトルμ_k（ｓ）のｉ番目のコンポーネントを、Σ_k（ｓ）（ｉ，ｉ）は、分散マトリクスΣ_k（ｓ）の、第ｉ行第ｉ列のコンポーネントを、それぞれ表す。そして、ｋクラスモデルの出力確率は、これらによって規定される。

【００４０】

なお、HMMは、上述したように、初期状態確率$\pi_k(q_h)$、遷移確率$a_k(q_i, q_j)$、および出力確率$b_k(q_i)$（O）によって規定されるが、これらは、学習用の音声データから特徴ベクトルを算出し、その特徴ベクトルを用いて、予め求めることとする。

【００４１】

ここで、HMMとして、図５に示したものを用いる場合には、常に、最も左の状態q_1から遷移が始まるので、状態q_1に対応する初期状態確率だけが１とされ、他の状態に対応する初期状態確率はすべて０とされる。また、出力確率は、式（９），（１０）から明らかなように、Ψ（ｔ）（ｉ，ｉ）を０とすると、特徴ベクトルの分散を考慮しない場合の連続HMMにおける出力確率に一致する。

【００４２】

なお、HMMの学習方法としては、例えば、Baum-Welchの再推定法などが知られている。

【００４３】

図４に戻る。識別関数演算部２１－ｋ（ｋ＝１，２，・・・，Ｋ）は、ｋクラスモデルについて、あらかじめ学習により求められている初期状態確率$\pi_k(q_h)$、遷移確率$a_k(q_i, q_j)$、および出力確率$b_k(q_i)$（O）によって規定される式（８）の識別関数$G_k(Z)$を記憶しており、特徴抽出部２からの特徴分布パラメータＺを引数として、識別関数$G_k(Z)$を演算し、その関数値（上述した観測確率）$G_k(Z)$を、決定部２２に出力する。識別関数演算部２１－ｓは、無音音響モデル補正部７から供給される初期状態確率$\pi_s(q_h)$、遷移確率$a_s(q_i, q_j)$、および出力確率$b_s(q_i)$（O）によって規定される、式（８）の識別関数$G_k(Z)$と同様の識別関数$G_s(Z)$を記憶しており、特徴抽出部２からの特徴分布パラメータＺを引数として、識別関数$G_s(Z)$を演算し、その関数値（上述した観測確率）$G_s(Z)$を、決定部２２に出力する。

【００４４】

決定部２２では、識別関数演算部２１－１乃至２１－ｋ、および識別関数演算部２１－ｓそれぞれからの関数値$G_k(Z)$（ここでは、関数値$G_s(Z)$を含む

ものとする）に対して、例えば、次式（１１）に示す決定規則を用いて、特徴分布パラメータＺ、すなわち、入力された音声が属するクラス（音響モデル）が識別される。

【数１１】

$$C(Z)=C_k \text{ , if } G_k(Z)=\max_i\{G_i(Z)\} \qquad \cdots (11)$$

ただし、Ｃ（Ｚ）は、特徴分布パラメータＺが属するクラスを識別する識別操作（処理）を行う関数を表す。また、式（１１）の第２式の右辺におけるmaxは、それに続く関数値G_i（Ｚ）（ただし、ここでは、ｉ＝ｓ，１，２，・・・，Ｋ）の最大値を表す。

【００４５】

決定部２２は、式（１１）にしたがって、クラスを決定すると、それを、入力された音声の認識結果として出力する。

【００４６】

図１に戻る。無音音響モデル補正部７は、ノイズ観測区間抽出部３から入力されるノイズ観測区間Ｔｎの音声データ（環境ノイズ）に基づいて、音声認識部６に記憶されている無音音響モデルに対応する識別関数G_s（Ｚ）を生成して音声認識部６に供給する。

【００４７】

具体的には、無音音響モデル補正部７では、ノイズ観測区間抽出部３から入力されるノイズ観測区間Ｔｎの音声データ（環境ノイズ）のＭ個のフレームの各フレームについて、特徴ベクトルＸが観測され、それらの特徴分布が生成される。

【数１２】

$$\{F_1(X), F_2(X), \cdots, F_M(X)\} \qquad \cdots (12)$$

なお、特徴分布｛F_i（Ｘ），ｉ＝１，２，・・・，Ｍ｝は、確率密度関数(Probabilistic Density Function)であるので、以下、無音特徴分布PDFとも記述する。

【００４８】

次に、無音特徴分布PDFを、次式（１３）に従い、図７に示すように、無音音響モデルに対応する確率分布$F_S(X)$に写像する。

【数１３】

$$F_S(X) = V(F_1(X), F_2(X), \cdots, F_M(X)) \quad \cdots (13)$$

ただし、Ｖは無音特徴分布PDF｛$F_i(X)$，ｉ＝１，２，・・・，Ｍ｝を無音音響モデル$F_S(X)$に写像する補正関数（写像関数）である。

【００４９】

この写像は、無音特徴分布PDFの記述によって様々な方法が考えられる。例えば、

【数１４】

$$F_S(X) = \sum_{i=1}^{M} \beta_i(F_1(X),F_2(X),\cdots,F_M(X),M) \cdot F_i(X) \quad \cdots (14)$$

$$= \sum_{i=1}^{M} \beta_i \cdot F_i(X) \quad \cdots (15)$$

ただし、$\beta_i(F_1(X), F_2(X), \cdots, F_M(X), M)$は、各無音特徴分布の重み関数であり、以下、$\beta_i$と記述する。なお、重み関数$\beta_i$は、次式（１６）の条件を満足するものである。

【数１５】

$$\sum_{i=1}^{M} \beta_i(F_1(X),F_2(X),\cdots,F_M(X),M) = \sum_{i=1}^{M} \beta_i \equiv 1 \quad \cdots (16)$$

【００５０】

ここで、無音音響モデルの確率分布$F_S(X)$が正規分布であると仮定し、また、各フレームの特徴ベクトルを構成するコンポーネントが無相関であると仮定すれば、無音特徴分布PDF｛$F_i(X)$，ｉ＝１，２，・・・，Ｍ｝の共分散行列Σ_iは対角線行列となる。ただし、この仮定の前提条件は、無音音響モデルの共分散行列も対角線行列であることである。したがって、各フレームの特徴ベクト

ルを構成するコンポーネントが無相関であれば、無音特徴分布PDF｛F_i（X），i＝1，2，・・・，M｝は、各コンポーネントに対応する平均と分散を持つ正規分布G（E_i，Σ_i）となる。E_iはF_i（X）の期待値であり、Σ_iはF_i（X）の共分散行列である。

【００５１】

さらに、ノイズ観測区間ＴｎのM個のフレームに対応する無音特徴分布の平均をμ_i、分散をσ_i^2と表すことにすれば、無音特徴分布の確率密度関数は、正規分布G（μ_i，σ_i^2）（i＝1，2，・・・，M）と表すことができる。したがって、各フレームに対応する平均μ_i、および分散σ_i^2を用い、以下に示す様々な方法によって演算される無音音響モデルの正規分布G（μ_s，σ_s^2）（上述したG_s（Z）に相当する）は、図７に示した無音音響モデルF_s（X）の近似分布となる。

【００５２】

無音音響モデルの正規分布G（μ_s，σ_s^2）を演算する第１の方法は、無音特徴分布｛G（μ_i，σ_i^2），i＝1，2，・・・，M｝を用い、次式（１７）に示すように、全てのμ_iの平均を無音音響モデルの平均値μ_sとし、次式（１８）に示すように、全てのσ_i^2の平均を無音音響モデルの分散σ_i^2とする方法である。

【数１６】

$$\mu_s = \frac{a}{M}\sum_{i=1}^{M}\mu_i \qquad \cdots (17)$$

$$\sigma_s^2 = \frac{b}{M}\sum_{i=1}^{M}\sigma_i^2 \qquad \cdots (18)$$

ここで、ａおよびｂは、シミュレーションにより最適な値が決定される係数である。

【００５３】

無音音響モデルの正規分布G（μ_s，σ_s^2）を演算する第２の方法は、無音特徴分布｛G（μ_i，σ_i^2），i＝1，2，・・・，M｝の期待値μ_iのだけを用い

、次式（１９），（２０）に従って、無音音響モデルの平均値μ_sと、分散σ_i^2を演算する方法である。

【数１７】

$$\mu_s = \frac{a}{M} \cdot \sum_{i=1}^{M} \mu_i \qquad \cdots (19)$$

$$\sigma_s^2 = b \cdot \left(\frac{1}{M} \cdot \sum_{i=1}^{M} \mu_i^2 - \mu_s^2 \right) \qquad \cdots (20)$$

ここで、ａおよびｂは、シミュレーションにより最適な値が決定される係数である。

【００５４】

無音音響モデルの正規分布Ｇ（μ_s，σ_s^2）を演算する第３の方法は、無音特徴分布｛Ｇ（μ_i，σ_i^2），ｉ＝１，２，・・・，Ｍ｝の組み合わせによって、無音音響モデルの平均値μ_sと、分散σ_s^2を演算する方法である。

【００５５】

この方法においては、各無音特徴分布Ｇ（μ_i，σ_i^2）の確率統計量をX_iとする。

【数１８】

$$\{X_1, X_2, \cdots, X_M\} \qquad \cdots (21)$$

【００５６】

ここで、無音音響モデルの正規分布Ｇ（μ_s，σ_s^2）の確率統計量をX_sとすれば、確率統計量X_sは、次式（２２）に示すように、確率統計量X_iと重み関数β_iの線形結合で表すことができる。なお、重み関数β_iは式（１６）の条件を満足している。

【数１９】

$$X_s = \sum_{i=1}^{M} \beta_i \cdot X_i \qquad \cdots (22)$$

【００５７】

そして、無音音響モデルの正規分布Ｇ（μ_s，σ_s^2）は、次式（２３）に示すように表される。

【数２０】

$$G(\mu_s, \sigma_s^2) = G(\sum_{i=1}^{M} \beta_i \mu_i, \quad \sum_{i=1}^{M} \beta_i^2 \sigma_i^2) \qquad \cdots (23)$$

【００５８】

なお、式（２３）を一般化するためには、重み関数β_iが、｛$\beta_i = 1/M$，ｉ＝１，２，・・・，Ｍ｝と仮定され、平均値μ_sと、分散σ_s^2には、係数が乗算される。

【数２１】

$$\mu_s = \frac{a}{M} \cdot \sum_{i=1}^{M} \mu_i \qquad \cdots (24)$$

$$\sigma_s^2 = \frac{b}{M^2} \cdot \sum_{i=1}^{M} \sigma_i^2 \qquad \cdots (25)$$

ここで、ａおよびｂは、シミュレーションにより最適な値が決定される係数である。

【００５９】

無音音響モデルの正規分布Ｇ（μ_s，σ_s^2）を演算する第４の方法では、無音特徴分布｛Ｇ（μ_i，σ_i^2），ｉ＝１，２，・・・，Ｍ｝の確率統計量X_iに対応する統計母集団$\Omega_i = \{f_{i,j}\}$を仮定する。ここで、

【数２２】

$$\{N_i \equiv N; \quad i = 1, 2, \cdots, M\}$$

とすれば、平均値μ_iは、次式（２６）によって得ることができ、分散σ_i^2は、次式（２８）によって得ることができる。

【数２３】

$$\mu_i = \frac{1}{N}\sum_{j=1}^{M} f_{i,j} \quad \cdots (26)$$

$$\sigma_i^2 = \frac{1}{N}\sum_{j=1}^{M} (f_{i,j}^2 - \mu_j)^2 \quad \cdots (27)$$

$$= \frac{1}{N}\sum_{j=1}^{M} f_{i,j}^2 - \mu_j^2 \quad \cdots (28)$$

【００６０】

式（２８）を変形すれば、次式（２９）の関係が成立する。

【数２４】

$$\frac{1}{N}\sum_{j=1}^{M} f_{i,j}^2 = \sigma_i^2 + \mu_i^2 \quad \cdots (29)$$

【００６１】

ここで、統計母集団の和Ω

【数２５】

$$\Omega = \bigcup_{i=1}^{M} \Omega_i$$

を考慮すれば、式（２６）から次式（３０），（３１）が導かれ、式（２９）から次式（３２）乃至（３１）が導かれる。

【数２６】

$$\mu_s = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} f_{i,j} \quad \cdots (30)$$

$$= \frac{1}{M}\sum_{i=1}^{M} \mu_i \quad \cdots (31)$$

$$\sigma_s^2 = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} (f_{i,j} - \mu_s)^2 \quad \cdots (32)$$

$$= \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} f_{i,j}^2 - \mu_s^2 \quad \cdots (33)$$

$$= \frac{1}{M}\sum_{i=1}^{M} (\sigma_i^2 + \mu_i^2) - \mu_s^2 \quad \cdots (34)$$

【００６２】

なお、実際には、式（３１）と式（３４）には係数が乗算されて用いられる。

【数２７】

$$\mu_s = \frac{a}{M}\sum_{i=1}^{M} \mu_i \quad \cdots (35)$$

$$\sigma_s^2 = b \cdot \left(\frac{1}{M}\sum_{i=1}^{M} (\sigma_i^2 + \mu_i^2) - \mu_s^2\right) \quad \cdots (36)$$

ここで、ａおよびｂは、シミュレーションにより最適な値が決定される係数である。

【００６３】

また、次式（２７）に示すように、分散σ_s^2に対してだけ、係数を乗算するようにしてもよい。

【数２８】

$$\sigma_s^2 = \frac{b}{M}\sum_{i=1}^{M} \sigma_i^2 + \frac{1}{M}\sum_{i=1}^{M} \mu_i^2 - \mu_s^2 \quad \cdots (37)$$

【００６４】

次に、音声認識装置の動作について説明する。

【００６５】

フレーム化部２には、マイクロフォン１で集音された音声データ（環境ノイズを含む認識対象の発話音声）が入力され、そこでは、音声データがフレーム化され、各フレームの音声データは、観測ベクトルａとして、ノイズ観測区間抽出部３、および特徴抽出部５に順次供給される。ノイズ観測区間抽出部３では、発話スイッチ４がオンとされたタイミングt_b以前のノイズ観測区間Ｔｎの音声データ（環境ノイズ）が抽出されて特徴抽出部５および無音音響モデル補正部７に供給される。

【００６６】

特徴抽出部５では、フレーム化部２からの観測ベクトルａとしての音声データが音響分析され、その特徴ベクトルｙが求められる。さらに、特徴抽出部５では、求められた特徴ベクトルｙに基づいて、特徴ベクトル空間における分布を表す特徴分布パラメータＺが算出され、音声認識部６に供給される。音声認識部６では、特徴抽出部５からの特徴分布パラメータを用いて、無音区間および所定数Ｋの単語それぞれに対応する音響モデルの識別関数の値が演算され、その関数値が最大である音響モデルが、音声の認識結果として出力される。また、音声認識部６では、無音音響モデル補正部７から入力される無音音響モデルに対応する識別関数を用いて、それまで記憶されていた無音音響モデルに対応する識別関数が更新される。

【００６７】

以上のように、観測ベクトルａとしての音声データが、その特徴量の空間である特徴ベクトル空間における分布を表す特徴分布パラメータＺに変換されるので、その特徴分布パラメータは、音声データに含まれるノイズの分布特性を考慮したものとなっており、また、無音区間を識別するための無音音響モデルに対応する識別関数が、発話直前のノイズ観測区間Ｔｎの音声データに基づいて更新されているので、音声認識率を大きく向上させることが可能となる。

【００６８】

次に、図８は、発話スイッチ４がオンとされてから発話が開始されるまでの無音区間Ｔｓを変化させたときの音声認識率の変化を測定した実験の結果を示している。

【００６９】

なお、図８において、曲線ａは無音音響モデルを補正しない従来の方法による結果を示しており、曲線ｂは第１の方法による結果を示しており、曲線ｃは第２の方法による結果を示しており、曲線ｄは第３の方法による結果を示しており、曲線ｅは、第４の方法による結果を示している。

【００７０】

実験の条件は、以下の通りである。認識される音声データは、高速道路を走行中の車内で集音されたものである。ノイズ観測区間Ｔｎは、２０フレームで約０．２秒である。無音区間Ｔｓは、０．０５秒、０．１秒、０．２秒、０．３秒、０．５秒とした。音声データの特徴抽出においては、MFCC(Mel-Frequency Cepstral Coefficients)ドメインで分析を実施した。認識の対象とする音声の発話者は、男女４人ずつ計８人であり、一人当たり３０３個の単語を離散して発話した。タスクは大語彙離散日本語で５０００ワードである。音響モデルは、HMMであり、良好な音声データを用いて予め学習されているものである。音声認識においては、Viterbiサーチ法でビーム幅を３０００とした。

【００７１】

なお、第１、第２、および第４の方法においては、係数ａを１．０とし、係数ｂを０．１とした。第３の方法においては、係数ａを１．０とし、係数ｂを１．０とした。

【００７２】

図８から明らかなように、従来の方法（曲線ａ）では、無音区間Ｔｓが長くなるのに伴って音声認識率が著しく低下しているが、本発明の第１乃至４の方法（曲線ｂ乃至ｅ）では、無音区間Ｔｓが長くなっても、音声認識率は、わずかしか低下しない。すなわち、本発明によれば、無音区間Ｔｓが変化しても、音声認識率はある程度のレベルを維持することが可能である。

【００７３】

以上、本発明を適用した音声認識装置について説明したが、このような音声認識装置は、例えば、音声入力可能なカーナビゲーション装置、その他各種の装置に適用可能である。

【００７４】

なお、本実施の形態では、ノイズの分布特性を考慮した特徴分布パラメータを求めるようにしたが、このノイズには、例えば、発話を行う環境下における外部からのノイズの他、例えば、電話回線その他の通信回線を介して送信されてくる音声の認識を行う場合には、その通信回線の特性なども含まれる。

【００７５】

また、本発明は、音声認識の他、画像認識その他のパターン認識を行う場合にも適用可能である。

【００７６】

なお、上記各処理を行うコンピュータプログラムは、磁気ディスク、CD-ROM等の情報記録媒体よりなる提供媒体のほか、インターネット、デジタル衛星などのネットワーク提供媒体を介してユーザに提供することができる。

【００７７】

【発明の効果】

以上のように、請求項１に記載のパターン認識装置、請求項６に記載のパターン認識方法、および請求項７に記載の提供媒体によれば、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルを生成し、記憶されている対応するものを更新するようにしたので、音声認識が開始された時から発話が開始される時までの時間が長くなるに伴って認識性能が低下することを抑止することが可能となる。

【図面の簡単な説明】

【図１】

本発明を適用した音声認識装置の構成例を示すブロック図である。

【図２】

図１のノイズ観測区間抽出部３の動作を説明するための図である。

【図３】

図１の特徴抽出部５の詳細な構成例を示すブロック図である。

【図４】

図１の音声認識部６の詳細な構成例を示すブロック図である。

【図５】

音声認識部６の動作を説明するための図である。

【図６】

図１の無音音響モデル補正部７の動作を説明するための図である。

【図７】

図１の無音音響モデル補正部７の動作を説明するための図である。

【図８】

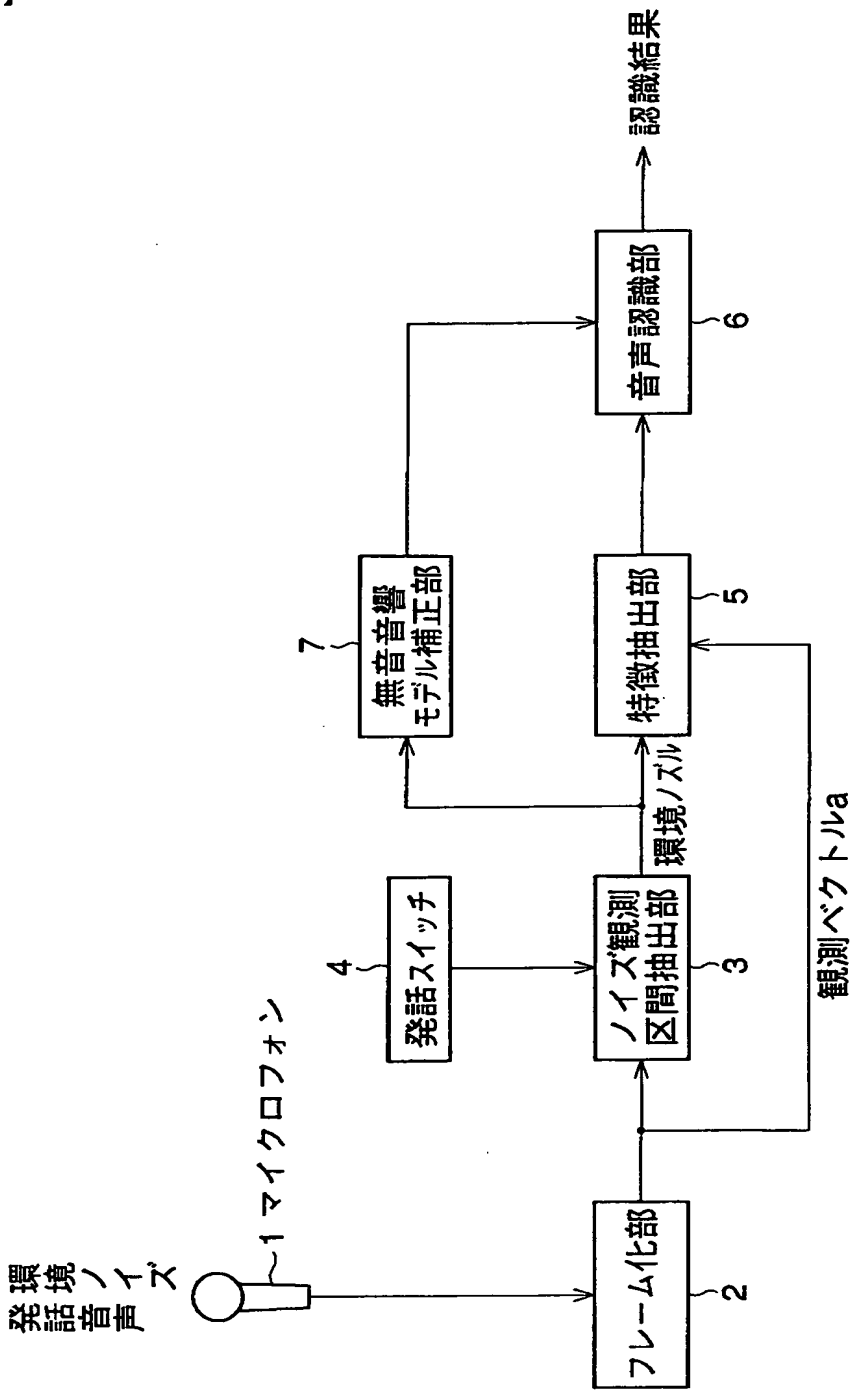
本発明を適用した音声認識装置の音声認識実験の結果を示す図である。

【符号の説明】

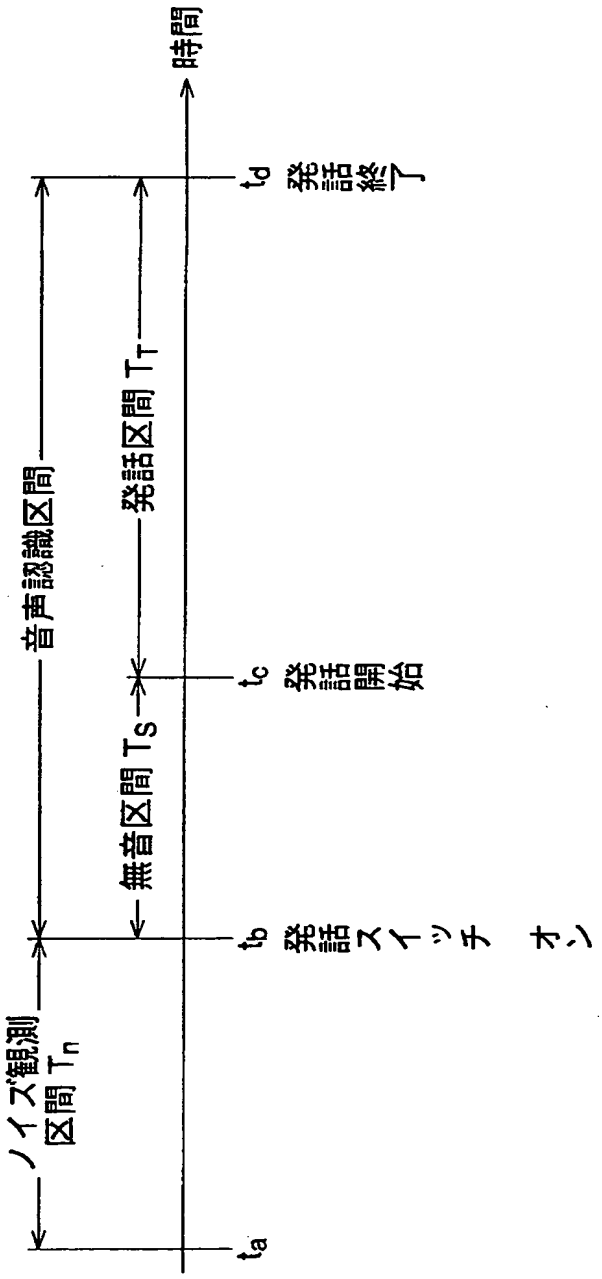
１　マイクロフォン，　２　フレーム化部，　３　ノイズ観測区間抽出部，４　発話スイッチ，　５　特徴抽出部，　６　音声認識部，　７　無音音響モデル補正部

【書類名】図面

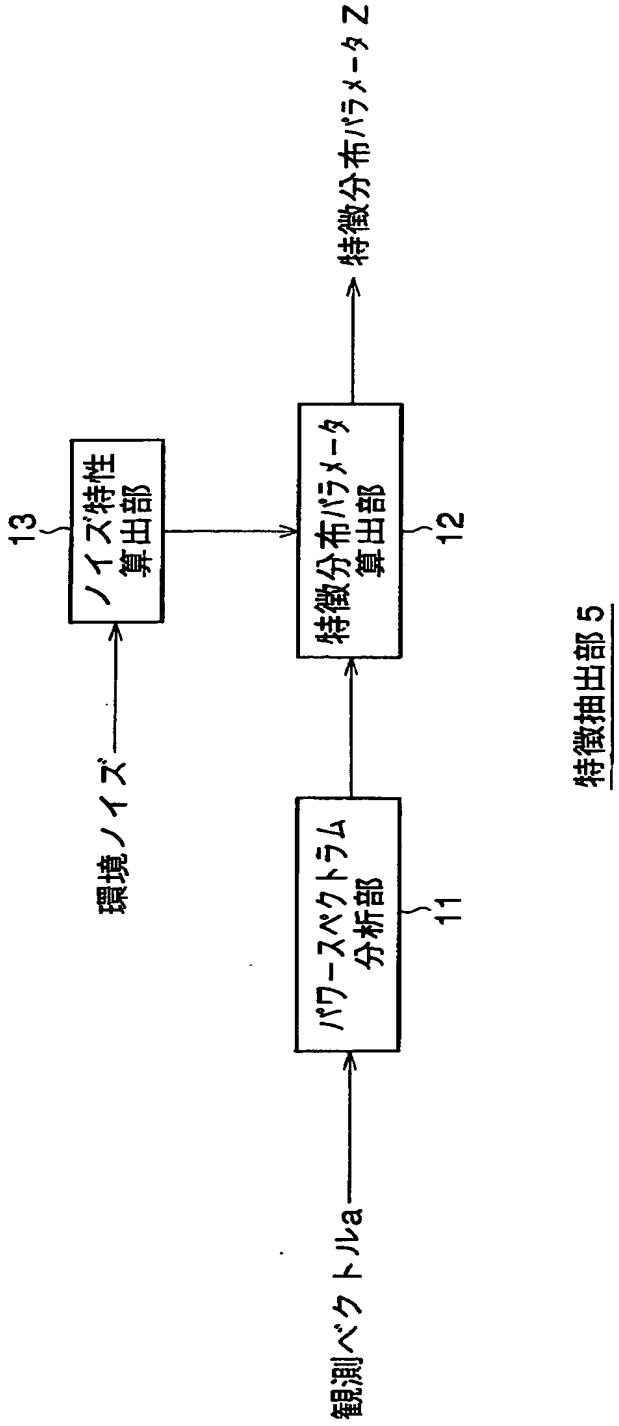
【図１】


認識結果
音声認識部
6
無音音響
モデル補正部
7
特徴抽出部
5
環境ノズル
発話スイッチ
4
ノイズ観測
区間抽出部
3
観測ベクトルa
マイクロフォン
1
環境ノイズ
発話音声
フレーム化部
2

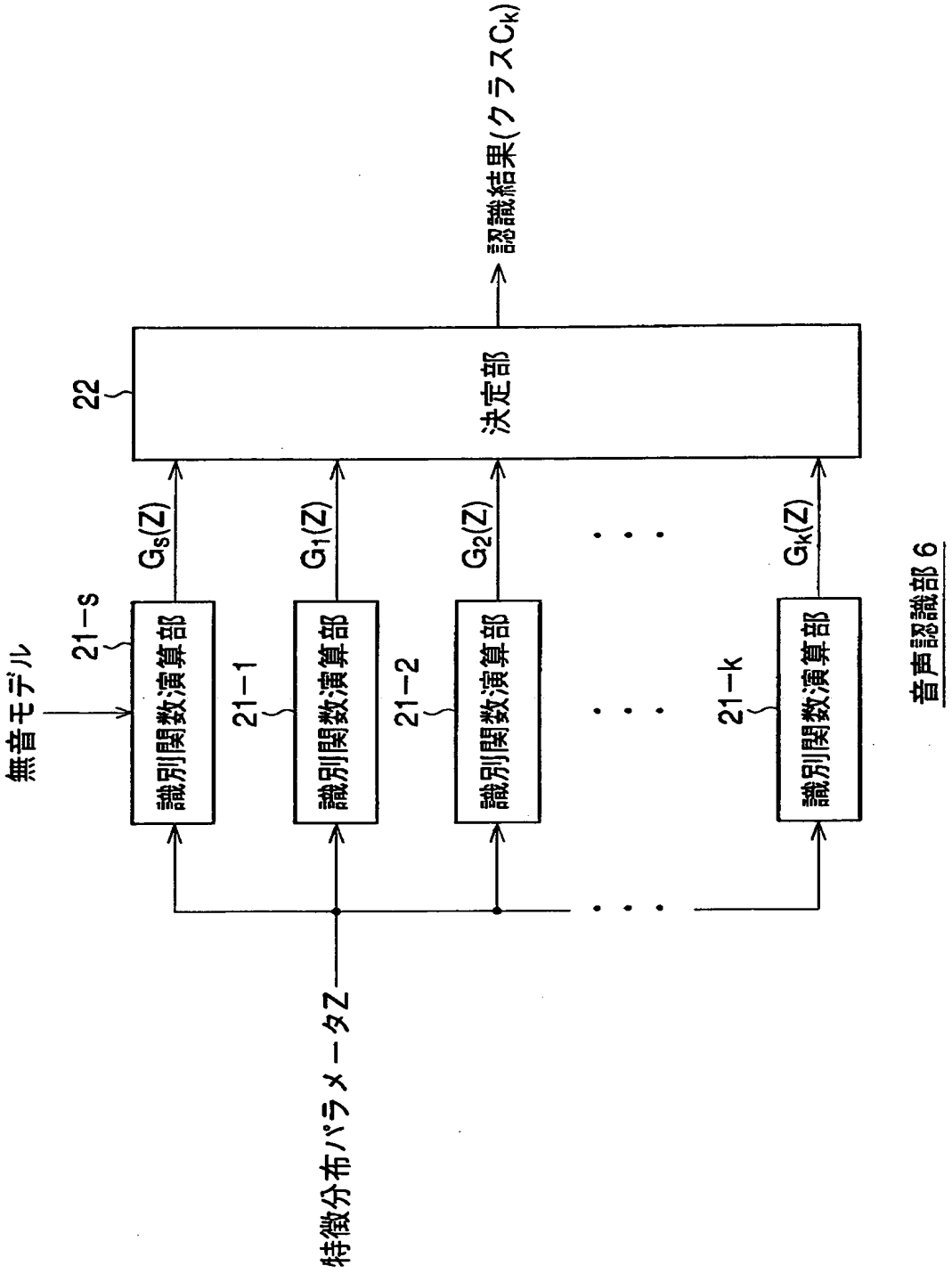
【図２】


時間
t_d 発話終了
t_c 発話開始
t_b 発話スイッチ　オン
t_a
音声認識区間
発話区間 T_T
無音区間 T_S
ノイズ観測区間 T_n

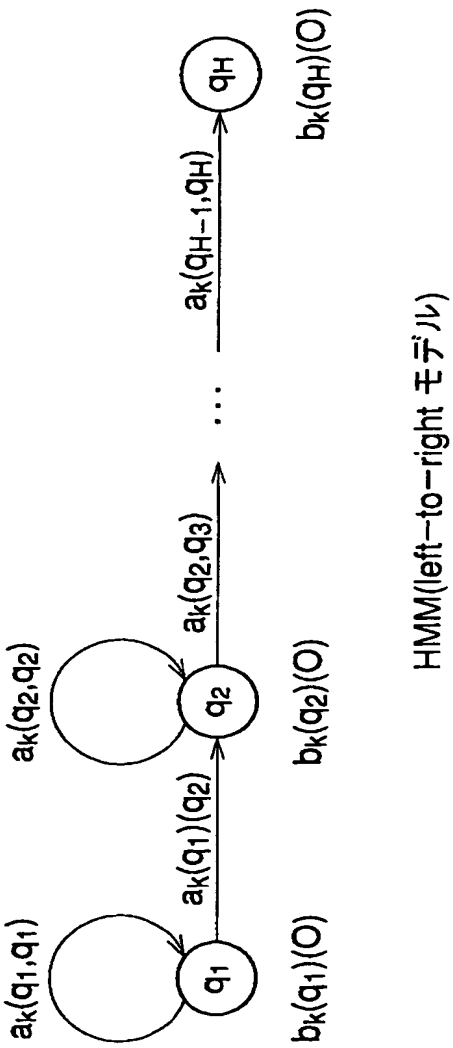
【図３】


観測ベクトルa
パワースペクトラム分析部
11
環境ノイズ
ノイズ特性算出部
13
特徴分布パラメータ算出部
12
特徴分布パラメータZ
特徴抽出部5

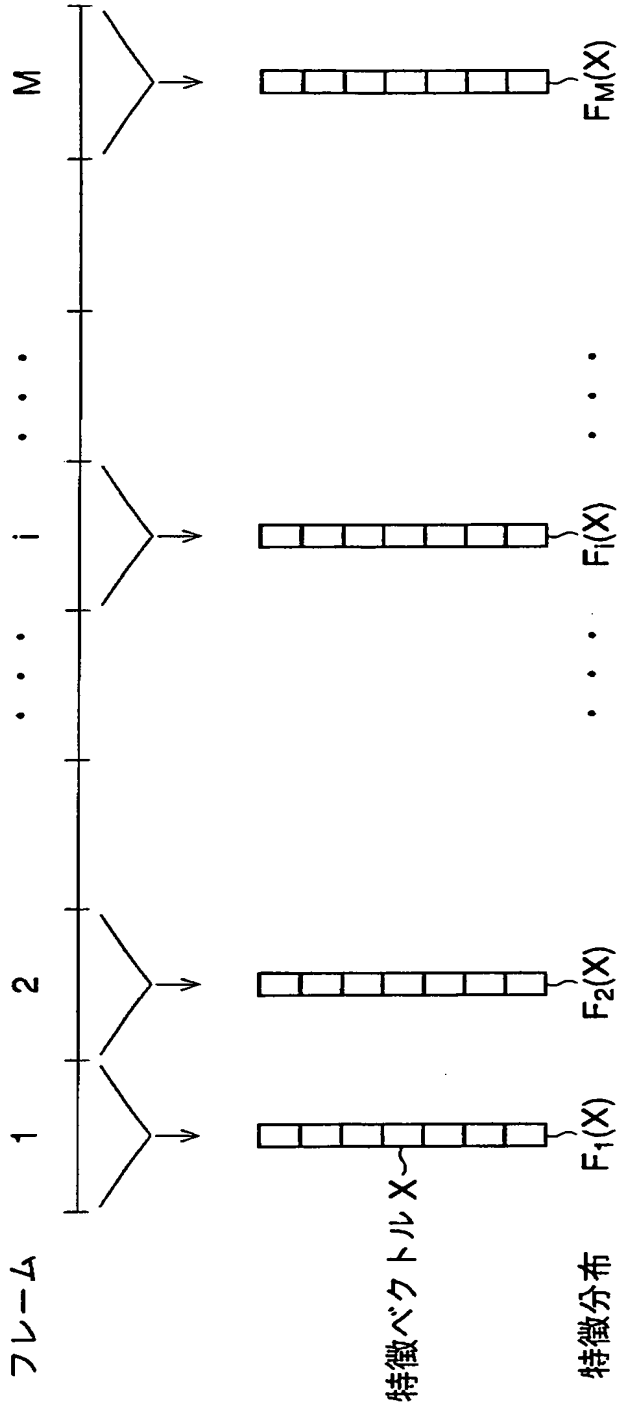
【図４】


認識結果(クラスCk)
決定部
22
Gs(Z)
G1(Z)
G2(Z)
Gk(Z)
無音モデル
21−s
21−1
21−2
21−k
識別関数演算部
特徴分布パラメータZ
音声認識部６

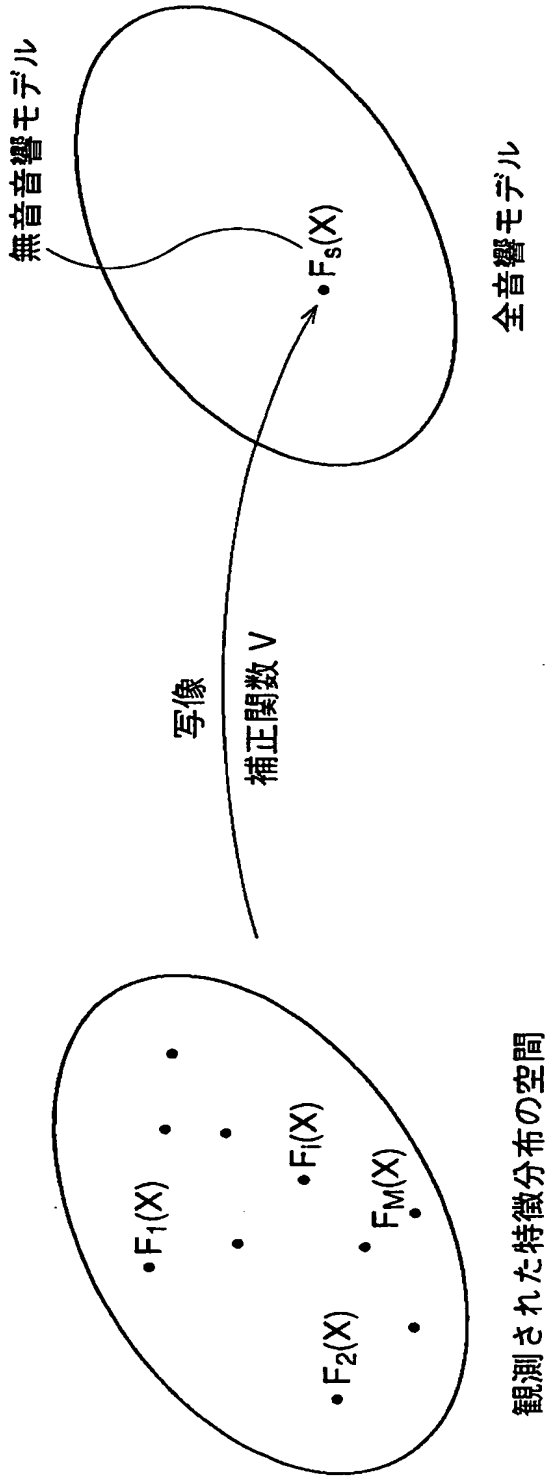
【図５】


q_1
q_2
q_H
$a_k(q_1,q_1)$
$a_k(q_1)(q_2)$
$a_k(q_2,q_2)$
$a_k(q_2,q_3)$
$a_k(q_{H-1},q_H)$
$b_k(q_1)(O)$
$b_k(q_2)(O)$
$b_k(q_H)(O)$
HMM(left-to-right モデル)

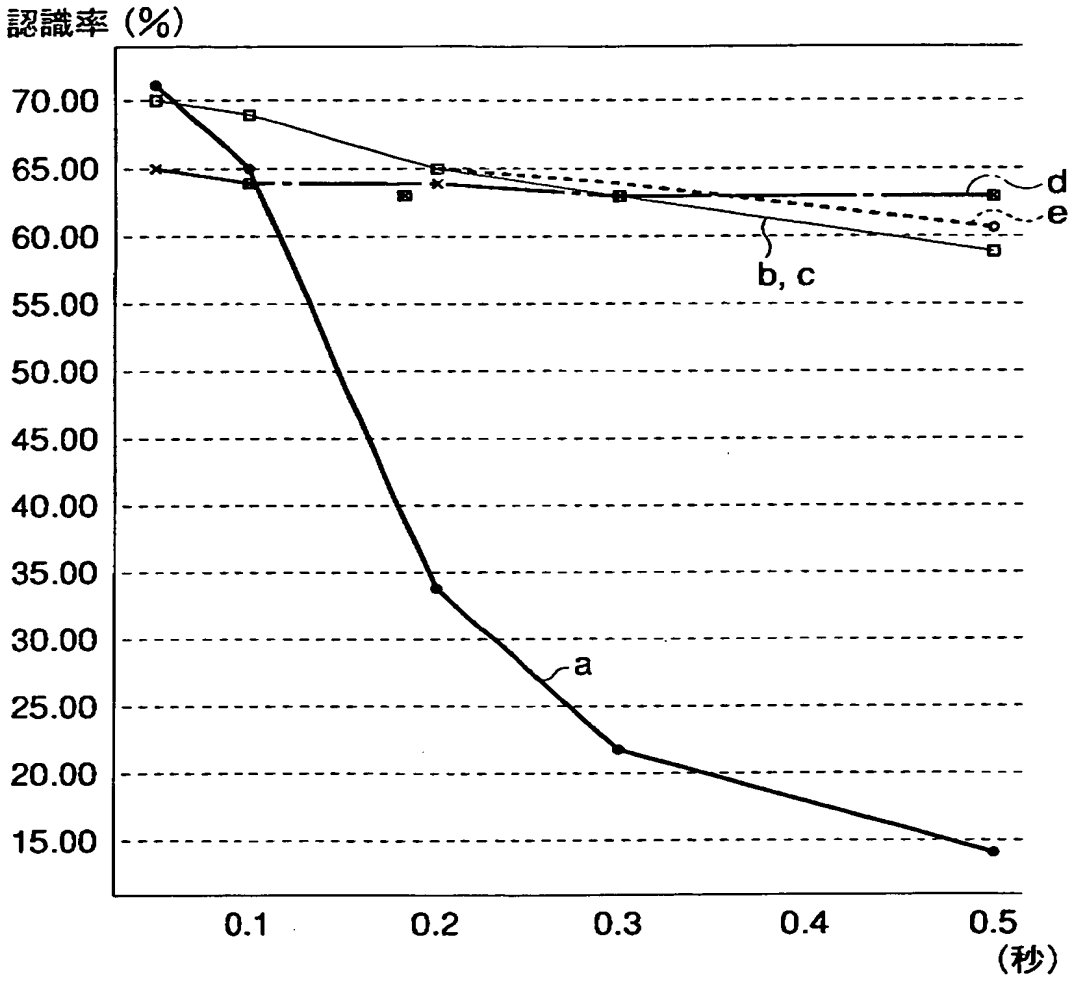
【図６】



【図７】



【図８】


認識率（%）
70.00
65.00
60.00
55.00
50.00
45.00
40.00
35.00
30.00
25.00
20.00
15.00
0.1
0.2
0.3
0.4
0.5
（秒）
a
b, c
d
e

【書類名】　　要約書

【要約】

【課題】　音声認識が開始された時から発話が開始される時までの時間が長くなるに伴って認識性能が低下することを抑止する。

【解決手段】　無音音響モデル補正部は、音声認識区間の初期に存在する無音区間Ｔｓを認識するために用いられる無音音響モデルを、ノイズ観測区間Ｔｎに含まれる環境ノイズに基づいて生成する。

【選択図】　　図２

出　願　人　履　歴　情　報

識別番号　　〔０００００２１８５〕

１．変更年月日　　１９９０年　８月３０日
［変更理由］　新規登録
住　所　東京都品川区北品川６丁目７番３５号
氏　名　ソニー株式会社